

Modelando o perfil do usuário para a construção de sistemas de recomendação: um estudo teórico e estado da arte

Fabrcio J. Barth

Resumo—O objetivo deste tutorial é descrever e sintetizar os conceitos e técnicas utilizados na concepção de sistemas de recomendação que permitem lidar com os perfis de usuários. O desenvolvimento de sistemas de recomendação que tratam os interesses do usuário pressupõe a resolução de dois subproblemas: (i) a criação e manutenção do perfil do usuário, e; (ii) o uso adequado desse modelo de perfil de usuário. Este trabalho é um tutorial teórico sobre o assunto. Trata-se de um texto voltado para pessoas que estão interessadas nos fundamentos teóricos sobre modelagem de perfil de usuários e sistemas de recomendação. O texto apresenta de forma ilustrativa esquemas que sintetizam os principais componentes utilizados na modelagem de perfis de usuários.

Index Terms—Sistemas de Recomendação, Modelagem de Usuário, Recuperação de Informação Adaptativa.

I. INTRODUÇÃO

Sistemas de Recomendação têm por objetivo recomendar itens (i.e., livros, músicas, artigos e fotos) que possam ser relevantes para o usuário. Nos últimos anos, algumas implementações de sistemas de recomendação têm adquirido uma considerável repercussão - Amazon.com¹ e Netflix² são alguns exemplos.

Estes sistemas de recomendação fazem uso de uma estrutura chamada perfil de usuário. Um perfil de usuário consiste, principalmente, de conhecimento sobre as preferências individuais que determinam o comportamento do usuário. Entende-se por preferências todas aquelas informações que são diretamente necessárias para a adaptação do comportamento do sistema aos interesses do usuário. Por exemplo, quantas vezes, qual a frequência e por quanto tempo um usuário acessa um determinado site na Web.

O objetivo deste tutorial é descrever e sintetizar os conceitos e técnicas utilizados na concepção de sistemas de recomendação que lidam com os interesses de usuários. O desenvolvimento de sistemas de recomendação que tratam os interesses do usuário pressupõe a resolução de dois subproblemas: (i) a criação e manutenção do perfil do usuário, e; (ii) o uso adequado desse modelo de usuário [1].

Nas próximas seções serão discutidas, com maior profundidade, questões relativas a representação, aquisição e manutenção de perfis de usuários. Este trabalho está estruturado da seguinte maneira: na seção II são apresentados as

principais definições relacionadas com o perfil do usuário; na seção III são descritas as técnicas normalmente empregadas para a criação e manutenção do perfil do usuário; na seção IV são descritas as principais abordagens para a utilização do perfil do usuário, e; na seção V são apresentadas algumas considerações finais.

II. PERFIL DO USUÁRIO

Há algum tempo é reconhecido que para se construir sistemas em que as pessoas e máquina cooperam para executar uma tarefa é importante adquirir algumas características significativas do usuário [2], [3]. Estas características são usadas para construir algum tipo de modelo ou perfil de usuário. Tradicionalmente, o modelo que é desenvolvido é um modelo de usuário típico daquele sistema, um modelo geral.

Não é necessário ir muito longe para imaginar um modelo geral. Basta pegar exemplos tradicionais de desenvolvimento de software. Por exemplo, a tarefa de análise de requisitos. Esta tarefa, em conjunto com outras, estabelece o que o usuário deseja do sistema.

Para o desenvolvimento de modelos gerais ou canônicos, tipicamente, são coletados dados sobre o desempenho médio das pessoas sobre várias tarefas em vários ambientes. Estes são utilizados no desenvolvimento de sistemas, por exemplo, no desenvolvimento de interfaces homem-máquina [4].

A maior fraqueza destes estudos é que eles assumem como premissa que todas as pessoas que estão envolvidas no processo pertencem a um conjunto homogêneo. Por trás desta premissa, os valores que caracterizam uma pessoa “típica” podem ser usados para desenvolver sistemas que podem ser usados por qualquer pessoa.

Um perfil ou modelo de usuário consiste, principalmente, de conhecimento sobre as preferências individuais que determinam o comportamento do usuário. Entende-se por preferências todas aquelas informações que são diretamente necessárias para a adaptação do comportamento do sistema aos interesses do usuário [5].

O termo perfil de usuário pode ser usado para descrever uma grande quantidade de conhecimento sobre usuários. Para tentar organizar melhor qual conhecimento e de que forma este conhecimento é armazenado, foi caracterizado um espaço tridimensional, com as seguintes dimensões [3]: (i) opção por um único modelo versus diversos modelos individuais

¹<http://www.amazon.com>

²<http://www.netflix.com>

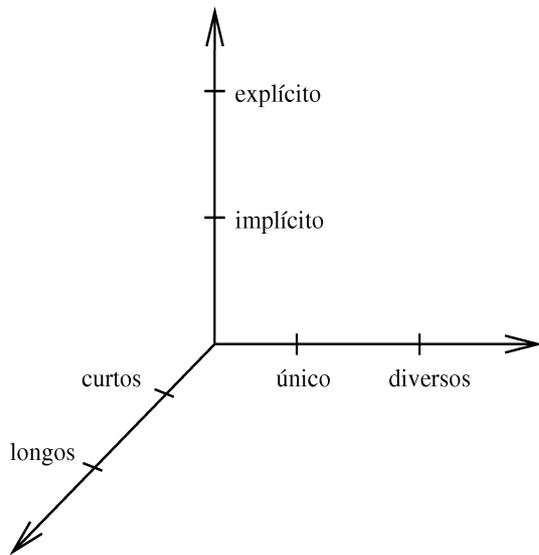


Figura 1. Espaço tridimensional do perfil de usuário (baseado em [1])

de diversos usuários; (ii) modelos especificados de maneira explícita versus modelos inferidos, e; (iii) modelo com conjunto de “*termos longos*”, características do usuário, tais como, o interesse e experiência do usuário versus modelo com conjunto de “*termos curtos*”, características do usuário, tais como, o problema que o usuário quer resolver neste momento. Na figura 1 é possível visualizar o espaço tridimensional mencionado acima.

Além disto, o perfil do usuário pode possuir informações pessoais a respeito do usuário, tais como: sua idade e profissão. Tais informações não são diretamente necessárias para a adaptação de um sistema ao usuário, mas podem ser usadas para categorizar o usuário em um estereótipo, que torna possível com que o sistema antecipe alguns dos comportamentos do usuário [3], [6].

A. Estereótipos

Estereótipos foram introduzidos por [2] como uma forma de organizar usuários de um sistema em grupos com o mesmo comportamento. As características sobre as pessoas, pelo menos grande parte, não são distribuídas aleatoriamente. Ou seja, existe um padrão de comportamento. Por exemplo, pessoas com maior poder aquisitivo viajam mais do que pessoas com menor poder aquisitivo.

O conhecimento sobre características comuns das pessoas pode ser inserido em uma coleção de estereótipos. Neste contexto, o conceito da palavra estereótipo está limitado à descrição de características que ocorrem em conjunto. A partir desta perspectiva, um estereótipo é um caminho simples para capturar algumas das estruturas que existem ao redor das pessoas. É uma forma que provê informações sobre pessoas similares [3].

Um estereótipo representa uma coleção de características relacionadas a um conjunto de predições sobre estas características. Por exemplo, a coleção de características pode ser: mulheres acima de 40 anos e com profissão igual a dona de casa. O conjunto de predições sobre estas características

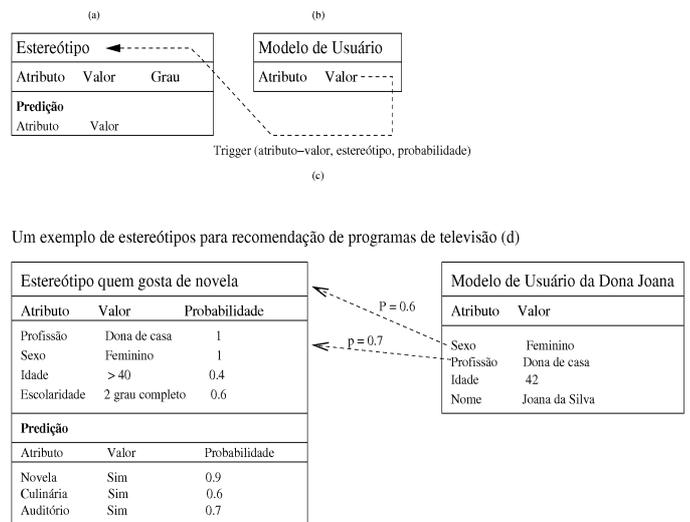


Figura 2. (a) Estrutura de um estereótipo; (b) Perfil de usuário; (c) Ação que causa a ativação de um estereótipo, e; (d) Exemplo de utilização

podem ser: gostam muito de novelas e gostam do programa da Ana Maria Braga³. Ambos os conjuntos podem ser representados como uma coleção de pares atributo-valor. Pode-se chamar cada atributo de faceta. O conjunto de predições representa conjecturas sobre aspectos relacionados a um grupo de pessoas. Na figura 2 é possível visualizar a estrutura do estereótipo e do perfil de usuário descrito acima.

Um estereótipo descreve apenas que uma coleção de características (aqui indicadas por atributos) geralmente ocorrem em conjunto, diferente de dizer que sempre ocorrem. Então, associada a cada faceta do estereótipo é necessário existir uma ordem de prioridade (*ranking*) que estima a probabilidade desta característica ser realmente parte do estereótipo.

No caso do perfil de usuários, alguns atributos podem ser facilmente observados. Eles servem como transações (*triggers*) (figura 2 item c) que causam a ativação de algum estereótipo. Cada transação é associada a uma probabilidade que referencia um determinado estereótipo aquela característica (atributo).

Todas as técnicas que tentam modelar o usuário de maneira automática fazem suposições sobre o mesmo. Suposições criadas pelo sistema a partir das informações sobre simples interações entre o usuário e o sistema. Como resultado, a possibilidade de erro deve ser sempre considerada. Para resolver isto, o sistema deve anexar justificativas para cada item que considera ser válido e nunca manter o perfil do usuário estático.

Na próxima seção serão vistos os métodos utilizados para criação e manutenção do perfil do usuário e os desafios inerentes a estas tarefas.

III. CRIAÇÃO E MANUTENÇÃO DO PERFIL DO USUÁRIO

Existem algumas premissas para o desenvolvimento de sistemas de recomendação que fazem uso de um perfil de usuário:

³Programa de variedades exibido em um rede de televisão brasileira.

- um sistema de recomendação que faz uso de um perfil de usuário não pode iniciar as suas atividades sem a criação de um perfil do usuário;
- é necessário representar o perfil do usuário (escolher uma técnica de representação);
- tais sistemas precisam de técnicas adequadas para gerar um perfil do usuário inicial, e;
- quando os usuários interagem com o sistema, eles fornecem informações sobre eles mesmos e sobre as suas atividades.

Com base nas premissas acima, é possível determinar cinco decisões de projeto que devem ser tomadas para o desenvolvimento de módulos que permitem a criação e manutenção do perfil do usuário [1] (figura 3): (i) a técnica de representação do perfil do usuário; (ii) a técnica utilizada para criação do perfil do usuário inicial; (iii) a técnica de aprendizado do perfil do usuário; (iv) a fonte de realimentação relevante que representa os interesses do usuário, e; (v) a técnica de adaptação do perfil do usuário.

Muitos métodos de aquisição do perfil de usuários tem sido apresentados na literatura. A utilização destes métodos depende da definição do perfil de usuário e da representação utilizados. Em geral, os métodos de aquisição do perfil de usuário podem ser agrupados em duas famílias: método explícitos e implícitos (não invasivos). Exemplo de um método explícito é a aplicação de questionários e um exemplo de método implícito é a visualização do comportamento do usuário e futura inferência das preferências de um usuário [6].

Técnicas de aprendizado de máquina (i.e., algoritmos de indução de árvores de decisão, algoritmos para segmentação - seção III-D) tem sido aplicadas em problemas de modelagem de usuários, principalmente para adquirir modelos individuais de usuários [7], [8], [9], [10] ou de grupos de usuários [11].

Observações sobre os comportamentos dos usuários podem fornecer um conjunto de exemplos de treinamento que um sistema de aprendizado de máquina pode usar para inferir um perfil de usuário [12]. Em tais situações, o uso de um sistema por um usuário é monitorado, os dados são coletados visando a construção do perfil do usuário. Por definição, o desenvolvimento de sistemas de recomendação pressupõe o desenvolvimento de módulos específicos, responsáveis por aprender os perfis de usuários através do monitoramento da interação do usuário com o sistema [13], [14], [15].

Características inerentes às aplicações de modelagem de usuário colocam inúmeros desafios às técnicas de aprendizado de máquina, incluindo: necessidade de grandes conjuntos de dados, necessidade de dados classificados, mudança no interesse dos usuários e complexidade computacional.

Este tutorial examina cada um desses desafios e descreve possíveis métodos para resolvê-los. Mas antes disso, são apresentados alguns aspectos históricos sobre modelagem de usuário e os métodos empregados nesta tarefa.

A. Histórico

A história sobre a aquisição de perfis de usuários automática passa por duas fases. As aplicações mais antigas que utilizavam a modelagem de usuários focavam na descrição dos

processos cognitivos por trás das ações dos usuários ou das diferenças entre as habilidades dos usuários e dos especialistas. Ou seja, a ênfase era colocada no desenvolvimento de modelos dos processos cognitivos do usuário. Nesta época, a área de modelagem de usuários era aplicada em Sistemas Tutores Inteligentes, por isso o interesse em questões cognitivas do usuário [16].

Numa fase posterior, passou-se a priorizar a modelagem do comportamento do usuário. Isso talvez, porque o campo de pesquisa tenha iniciado predominantemente com a modelagem de estudante e a demanda por comércio eletrônico e recuperação de informação na Web fizeram o campo apontar para outra direção. Passando a valorizar mais a aquisição das preferências e necessidades dos usuários.

Por fim, pode-se dizer que a importância dada ao campo de modelagem de usuários vem aumentando. Isto pode ser visto através dos vários *workshops* que têm como assunto principal a modelagem de usuários, dentre eles, pode-se citar: *Machine Learning for User Modeling* [17], [18] e *International Conference on User Modeling* [19].

As próximas seções apontam para os desafios do processo de aquisição ou identificação automática do modelo de usuário.

B. Técnicas para representação do perfil do usuário

As informações contidas no perfil do usuário devem ser relevantes para o sistema de recomendação que se deseja desenvolver. A técnica de representação usada na concepção do perfil deve ser adequada para representar as informações relevantes do contexto em que o sistema atua.

Diversas técnicas de representação de perfil do usuário tem sido usadas nos mais diversos tipos de sistemas. Por exemplo, alguns sistemas mantêm a lista de compras efetuadas, o histórico de navegação ou o conteúdo da caixa postal como o perfil do usuário. Estes modelos são chamados de baseados em histórico [20].

Outra maneira de representação do perfil do usuário é considerá-lo como um vetor. Por exemplo, em sistemas de filtragem de informação que fazem uso do perfil do usuário [21], o perfil pode ser definido como um vetor \vec{s}_j com as probabilidades que refletem o quanto uma determinada característica w_k é relevante para o usuário ($p(w_k | Relevante)$).

Outra maneira é utilizar formas de representação que são usadas em uma ampla variedade de tarefas de aprendizagem, como por exemplo, árvores de decisão [22]. Na figura 4 é possível visualizar uma árvore de decisão que representa as preferências do usuário correspondentes ao agendamento de reuniões.

A informação contida na árvore de decisão também pode ser representada por um conjunto de regras [23]. Por exemplo, a seguinte regra é obtida através de um dos ramos da árvore:

```
SE dia = sexta
E urgente = sim
E período = manhã
ENTÃO Sim
```

Esta regra significa: se o dia da semana for sexta-feira e a reunião é considerada urgente e o período proposto é na parte da manhã então o sistema poderá marcar a reunião.

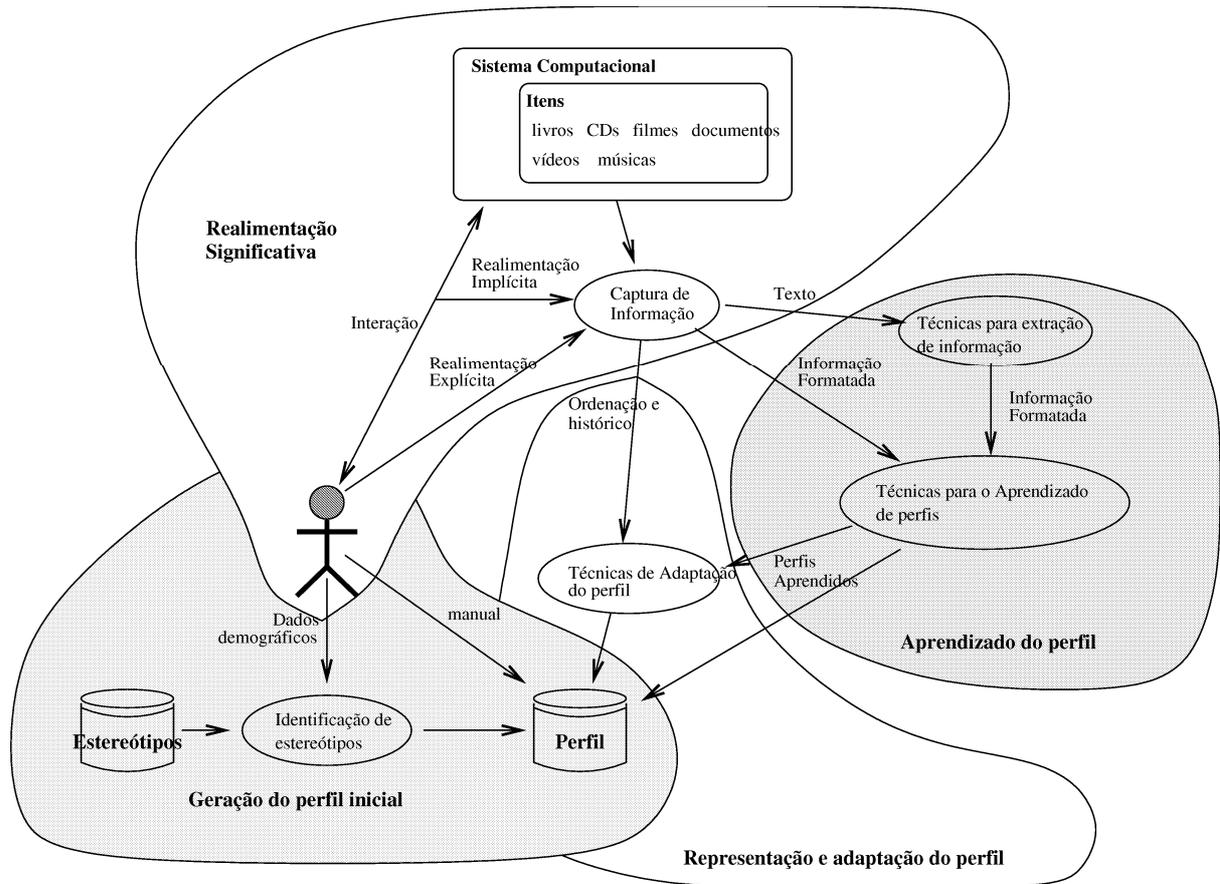


Figura 3. Criação e manutenção do perfil do usuário (baseado em [1])

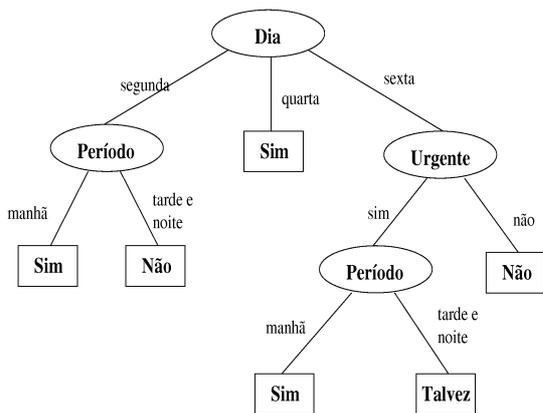


Figura 4. Exemplo de representação do perfil do usuário utilizando árvores de decisão

```

inClass(A):-
    author(A,'S. Franklin').
inClass(A):-
    author(A,'A. Graesser').
inClass(A):-
    titleword(A,agent),
    titleword(A,taxonomy).
inClass(A):-
    textword(A,agent),
    textword(A,learning),
    textword(A,reactive),
    textword(A,autonomous).
inClass(A):-
    type(A,'http://www.agents.edu/').
    
```

Figura 5. Perfil de usuário baseado em regras

Pode-se também optar por outras formas de representação, como por exemplo, através de regras em *Prolog*[24], subconjunto de cláusulas de primeira-ordem. Na figura 5 é possível visualizar um perfil de usuário baseado em regras deste tipo.

Estas regras representam condições para que um texto possa ser classificado como um texto sobre agentes, por exemplo.

C. Fontes para criação inicial do perfil do usuário

É desejável que o sistema de recomendação consiga o mais rápido possível informações sobre o usuário a ponto de fornecer adaptações úteis logo no início da interação do usuário com o sistema computacional. Este requisito impõem diversas dificuldades na concepção de um sistema de recomendação. Por exemplo, usualmente os usuários não perdem muito tempo definindo os seus interesses para criar o seu perfil.

Por essas razões, iniciar e manter perfis de usuários é um aspecto difícil no desenvolvimento de sistemas de

recomendação. Segundo [1], o nível de automação da aquisição de perfis de usuários pode partir da entrada manual de dados, para procedimentos semi-automáticos (estereótipos e conjunto de exemplos de treinamento) até o reconhecimento automático de perfis.

As técnicas mais comumente utilizadas são [1], [25]:

- *perfil inicial vazio*: neste caso não existe fase inicial. O perfil do usuário é populado através de um mecanismo automático quando o usuário começa a interação com o sistema;
- *manual*: neste método, um sistema com abordagem questiona os usuários para que eles registrem os seus interesses em um formulário de palavras chaves, tópicos, etc. Alguns problemas com este método podem ser identificados, por exemplo, este método exige muito esforço por parte do usuário e em alguns momentos o usuário não sabe descrever ou não sabe quais são os seus efetivos interesses;
- *estereótipos*: a criação de um perfil inicial usando o conceito de estereótipo é baseado no fato de que criar um modelo inicial é na verdade, um problema de classificação que permite gerar previsões iniciais sobre o usuário [5]. Tipicamente, os dados utilizados neste processo são dados pessoais (nome, endereço, idade, sexo, etc.), dados geográficos (CEP, cidade, estado, etc.), entre outros, e;
- *conjunto de treinamento*: o conjunto de treinamento é uma coleção de exemplos de interação do usuário com o sistema computacional que é usado para inferir um perfil inicial do usuário.

D. Técnicas para aprendizado do perfil do usuário

Na seção anterior foram descritas algumas fontes de informação potencial para a construção do perfil do usuário. As técnicas para aprendizado do perfil do usuário fazem uso destas fontes de informação para inferir o perfil do usuário.

Alguns sistemas de recomendação não utilizam nenhuma técnica para aprendizado do perfil do usuário. Entre estes, estão aqueles que fazem uso apenas de estereótipos. No entanto, a grande maioria utiliza alguma forma para o aprendizado do perfil do usuário. As técnicas para aprendizado podem ser classificadas em três: técnicas de extração de informação estruturada, agrupamento⁴ e classificação (figura 6).

Quando a fonte de informação não é estruturada (i.e., texto), algum tipo de pré-processamento é necessário para extrair informação estruturada relevante (figura 7). Tipicamente, este processo é composto de duas fases: seleção e indexação da informação. A seleção de palavras-chave utiliza diferentes abordagens para reduzir o número de palavras: *stop-words* e *stemming* [26]. As técnicas de indexação da informação utilizam formas para calcular a relevância de uma palavra-chave, como por exemplo, a técnica conhecida por TF-IDF (*Term Frequency / Inverse Document Frequency*) ou, simplesmente, o cálculo da frequência das palavras [27]. Nestes casos, o documento é representado como um vetor de palavras-chave

com os seus respectivos pesos. Nesta abordagem, por exemplo, as técnicas para recomendação de documentos mais utilizadas são aquelas baseadas na regra de *Bayes* [1].

A idéia básica das técnicas de agrupamento é organizar usuários similares em grupos, baseado nos dados disponíveis. Por exemplo, na implementação do sistema ACR News [28], as transações dos usuários, as URL navegadas, são mapeadas em um espaço multi-dimensional. Este espaço é particionado em agrupamentos representando os grupos que possuem transações similares. Finalmente, os agrupamentos são verificados com uma sessão ativa de um usuário para recomendar URL interessantes para o usuário. Métodos colaborativos tradicionais (seção IV-B) geralmente calculam a similaridade entre um objeto e um conjunto de agrupamentos.

Como técnicas de classificação podem ser utilizados algoritmos de indução de árvores de decisão, tais como ID3 [29] e C4.5 [22], para perfis de usuários que utilizam árvores de decisão como técnica de representação (figura 4). Pode-se utilizar programação lógica indutiva (*Inductive Logic Programming - ILP*) [30] para gerar regras que representam os interesses do usuário (figura 5) [28].

É importante salientar que um algoritmo de aprendizagem necessita de uma amostragem adequada para treinamento e indução visando possibilitar a construção de modelos mais precisos. Este problema pode ser tratado de várias maneiras:

- utilizar abordagens de aprendizado onde se cria um novo modelo a partir da modificação de um modelo inicial comum [31], [32];
- utilizar abordagens de aprendizado que são razoavelmente precisas com poucos exemplos, se os novos exemplos são parecidos com os exemplos de treinamento (por exemplo, os algoritmos do vizinho mais próximo) [32], e;
- em alguns casos, é possível estruturar a tarefa de forma que o modelo aprendido não precisa reproduzir exatamente as necessidades, decisões e desejos do usuário - o modelo pode apenas indicar algumas opções e deixar a decisão final para o usuário, ou seja, pode-se optar por modelos menos precisos [33].

Abordagens de aprendizado de máquina supervisionadas requerem dados classificados. Porém, a classificação de exemplos extraídos a partir da simples observação do comportamento do usuário [34] pode não corresponder a realidade. Uma solução seria requerer alguma ação adicional de classificação por parte do usuário, o que pode muitas vezes não funcionar.

Uma outra solução seria usar um conjunto inicial pequeno de exemplos classificados para inferir a classificação para um conjunto maior de exemplos, que será usado para treinar o algoritmo de aprendizado [35], [36].

E. Formas de realimentação

Os interesses de um usuário mudam ao longo do tempo. Um sistema de recomendação precisa de informação atualizada sobre os interesses do usuário para manter o perfil sempre de acordo com as necessidades do usuário. As duas maneiras mais comuns para obter realimentação relevante é através da informação fornecida pelo usuário de maneira explícita

⁴do inglês, *clustering*

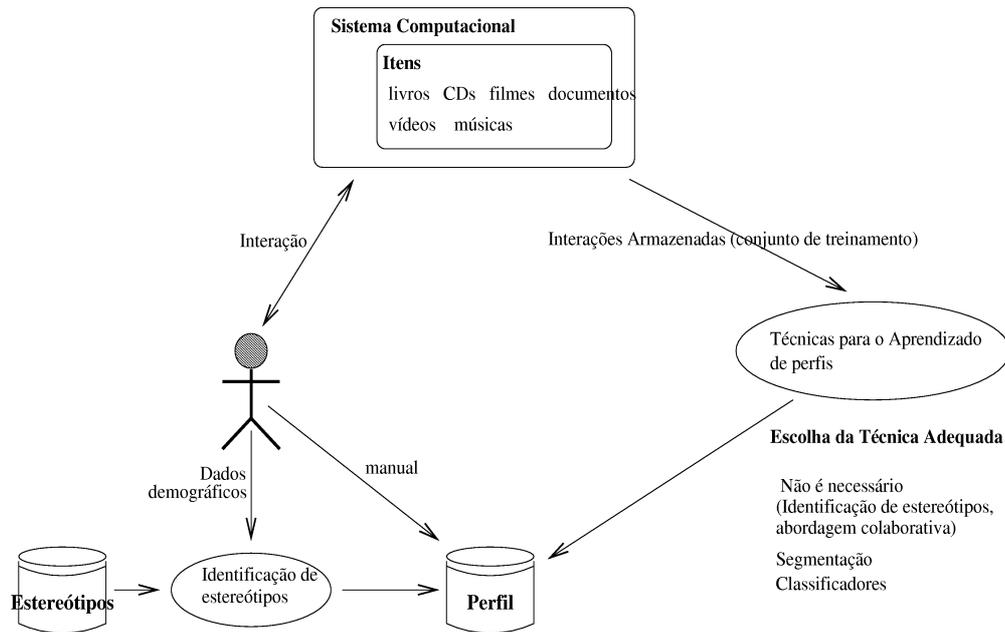


Figura 6. Fontes para criação inicial do perfil do usuário e técnicas para o aprendizado do perfil do usuário (baseado em [1])

ou implícita. Pode-se dividir as formas de realimentação em quatro [1] (figura 8):

- *sem realimentação*: para aqueles sistemas de recomendação que não precisam alterar o perfil do usuário a realimentação do usuário não é relevante;
- *realimentação explícita*: o usuário pode indicar a sua opinião sobre um determinado objeto recomendado ou sugerido pelo sistema de recomendação. A realimentação explícita que o usuário pode fornecer é dividido em *gosto/não gosto*⁵, ranqueamento, baseado na escala de Linkert e comentários textuais. Na forma *gosto/não gosto* o usuário apenas informa se o objeto ou a indicação é útil ou não [37], [32]. O usuário também poderá ordená-los de acordo com a relevância [38] ou inserir comentários de maneira textual como acontece no site *Grouplens* [39]. É claro que esta última forma de realimentação é mais rica e difícil de ser tratada. A realimentação explícita tem a grande vantagem da simplicidade. No entanto, em aplicações práticas, a realimentação explícita possui sérios problemas: (i) escalas numéricas não são adequadas para descrever a reação humana obtida a partir de uma experiência com um sistema de recomendação, e; (ii) raramente, os usuários fornecem realimentação (*cold-start problem* [40], [41]);
- *realimentação implícita*: significa que o sistema automaticamente irá inferir as preferências do usuário através do monitoramento das suas ações. A maioria dos métodos de realimentação implícita analisam as páginas visitadas [34], a lista de itens comprados [20], o tempo gasto com a leitura de um documento e se o usuário imprimiu ou salvou o documento.
- *abordagens híbridas*: como é comum em muitas tecnologias, os sistemas com o melhor desempenho são

aqueles que combinam diversas tecnologias existentes. Nesta área, a realimentação implícita pode ser combinado com a explícita. As técnicas de realimentação implícita servem para diminuir as responsabilidades do usuário, enquanto que as técnicas de realimentação explícita fornecem informação adicional para melhorar a acurácia do sistema.

As primeiras abordagens do uso de aprendizado de máquina para a modelagem de usuários frequentemente desenvolviam novas técnicas de aprendizado de máquina, geralmente *ad hoc*, para um determinado fim. Mais recentemente a tendência tem sido trabalhar mais em uma representação adequada do problema na forma de exemplos de treinamento e correspondente classificação dos exemplos de forma a utilizar algoritmos conhecidos na vasta literatura disponível sobre aprendizagem de máquina e classificação.

Um defeito em potencial dessa abordagem seria levar a soluções que não são especificamente talhadas para as características da modelagem de usuários, que é conhecida como uma tarefa muito dinâmica na medida em que os atributos que caracterizam um usuário mudam ao longo do tempo. E, para dar conta disso, é importante que os algoritmos sejam capazes de ajustar a essas mudanças rapidamente, problema que é caracterizado como uma mudança nos interesses do usuário ao longo do tempo (*concept drift* [42], *interest drift* [43]).

F. Formas de adaptação do perfil

Com o objetivo de enfatizar a importância deste problema e ilustrar melhor os desafios envolvidos, é descrito nesta seção o desenvolvimento de perfis de usuário para aplicações de recuperação de informação. O principal objetivo é aprender modelos de interesse ou necessidade de usuários com o intuito de facilitar a recuperação de informação relevante.

⁵do inglês, *like/dislike*

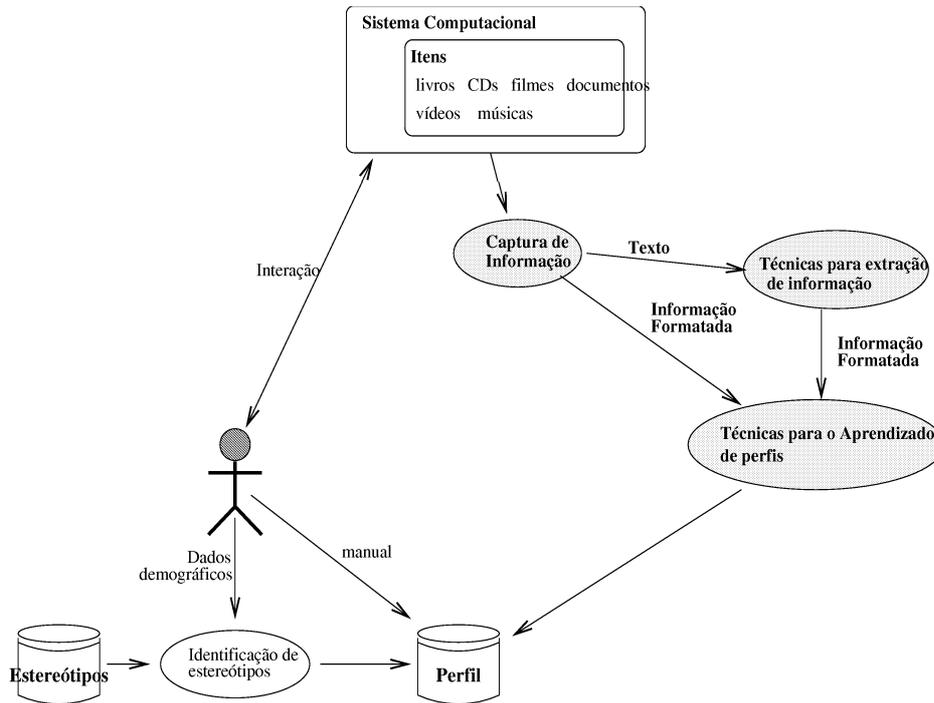


Figura 7. Captura de informação para o aprendizado do perfil do usuário (baseado em [1])

A maioria dos trabalhos sobre filtragem e recuperação de informação baseada em conteúdo tenta adquirir automaticamente os perfis de usuários como sendo uma tarefa de classificação de textos (por exemplo, [44], [45] e [46]). Nestes sistemas, o conjunto de documentos (texto) é ordenado pelo usuário (i.e., documentos interessantes versus não interessantes) e usados como entrada para um algoritmo de aprendizado. A classificação resultante pode ser interpretada como um modelo dos interesses do usuário induzido automaticamente.

Neste caso, uma hipótese razoável seria que quanto mais dados de treinamento melhor seria a recomendação. Por outro lado, os interesses dos usuários mudam ao longo do tempo e um classificador que se baseia num grande conjunto de treinamento que reflete interesses passados pode ter desempenho pior do que um limitado a dados recentes, mas que reflete os interesses atuais do usuário. Este exemplo ilustra que um bom algoritmo classificador de textos não necessariamente é um algoritmo útil de modelagem de usuários.

Algumas soluções iniciais para este problema foram emergindo na literatura, a idéia central está baseada em limitar o conjunto de treinamento à uma janela de tempo, onde o tamanho da janela depende dos indicadores observados, tais como: mudanças repentinas na distribuição dos dados [47], [42].

Segundo [48], o esquecimento gradual também é outra opção de solução que trabalha com a idéia de que as observações mais recentes devem ser mais importantes para o algoritmo de aprendizado do que as antigas e que a importância de uma observação deve decrescer com o tempo [43].

Restringir a criação dos perfis apenas ao uso de dados

recentes pode levar a simplificar demais os modelos. Isto é, estes classificariam com alta precisão as instâncias que são similares a dados colhidos recentemente, mas teriam um desempenho pobre sobre outras instâncias. A solução para esse problema é usar um modelo dual que classifica as instâncias, primeiro consultando um modelo treinado com dados recentes e, em seguida, delegando a classificação para um modelo treinado por mais tempo, caso o modelo recente não fizer uma previsão com segurança suficiente [49], [32], [12].

Segundo [1], as formas de adaptação do perfil do usuário podem ser classificadas em (figura 9): (i) *manual*: o usuário pode mudar o seu perfil quando for de interesse. É claro que os sistemas que optam por esta opção devem permitir a edição do perfil do usuário. Um aspecto negativo desta abordagem é que ela requer muito esforço por parte do usuário; (ii) *adição de novas informações*: simplesmente adiciona-se nova informação adquirida através da realimentação do usuário, e; (iii) *esquecimento gradual*: utiliza uma função de esquecimento.

G. Desafios na criação e manutenção do perfil do usuário

A criação e manutenção do perfil do usuário impõem alguns desafios. Em relação a representação do perfil do usuário, a tendência de novos projetos é utilizar técnicas de representação do perfil com maior poder de representação. No entanto, é evidente que a utilização de novas técnicas de representação implica na mudança da geração, utilização e adaptação do perfil do usuário. Todos estes detalhes devem ser cuidadosamente considerados no desenvolvimento de um sistema de recomendação.

Outra questão está relacionada a geração inicial do perfil do usuário. É tendência também criar técnicas que permitem, o

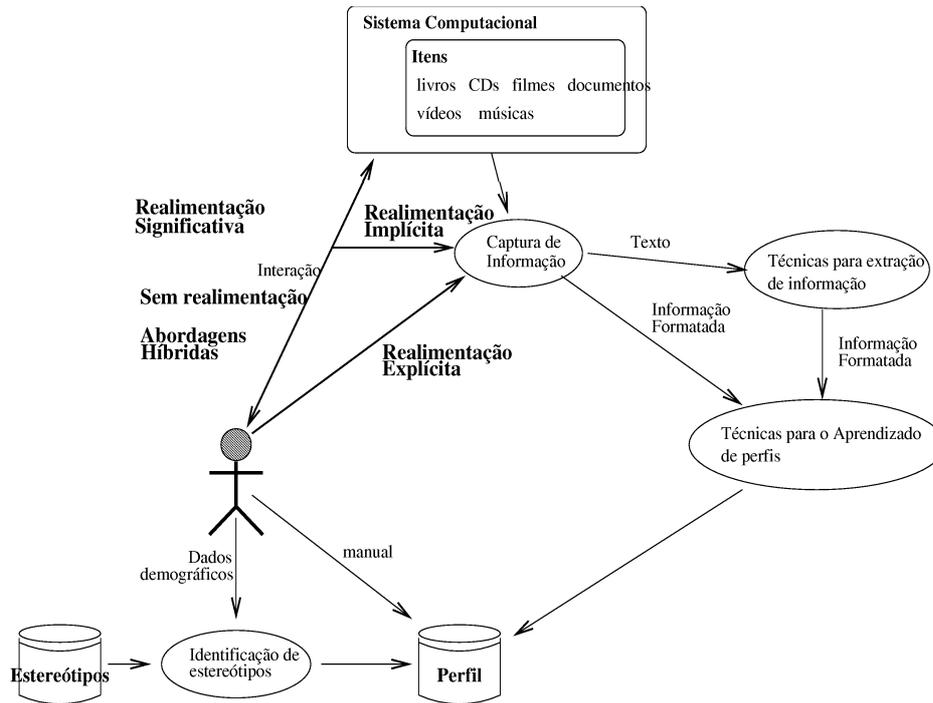


Figura 8. Formas de realimentação (baseado em [1])

mais rápido possível e com o mínimo de interações homem-máquina possíveis, criar um perfil do usuário inicial acurado.

Enquanto o crescimento da indústria da computação tem criado novas oportunidades para dar suporte aos usuários através do uso de modelos de usuário detalhados, a quantidade de informações disponíveis e o número de usuários que elevam a complexidade computacional é uma questão fundamental para a viabilidade da modelagem de usuários automática com a precisão exigida no campo da pesquisa em sistemas de larga escala.

Nesse sentido, no contexto dos sistemas usados em larga escala, um algoritmo menos preciso pode ser preferido a outro mais preciso se requerer menor tempo de processamento - ao contrário dos sistemas desenvolvidos em pequena escala, onde se privilegia a precisão.

As soluções apontadas para o uso de algoritmos de aprendizado (com elevada complexidade computacional) para a modelagem de usuários eficiente incluem, por exemplo, aplicar esses modelos onde eles podem ser aprendidos *off-line*, ou seja, sem as restrições de tempo real que requerem tempos de resposta reduzidos. Nesse caso haveria um pré-processamento que aceleraria o tempo de resposta do processo realizado *on-line* [12].

IV. UTILIZAÇÃO DO PERFIL DO USUÁRIO

Segundo [1] existem quatro principais abordagens para utilizar o perfil do usuário: filtragem demográfica, filtragem baseada em conteúdo, filtragem colaborativa e abordagem híbrida (figura 10). Filtragem demográfica usa descrições de pessoas para aprender a relação entre um item particular e o tipo de pessoa que gosta deste item.

Filtragem baseada em conteúdo usa descrições do conteúdo dos itens para aprender a relação entre um único usuário e diversos itens. Diversos métodos para cálculo da relevância entre perfil de usuários e itens são usados para comparar os interesses dos usuários e os itens analisados.

A filtragem colaborativa usa a realimentação fornecida por um conjunto de pessoas, que diz respeito a um conjunto de itens, para realizar recomendações, ignorando o conteúdo dos itens. Esta abordagem pode ser usada quando podemos assumir que um usuário se comporta de forma semelhante à outros usuários, ou quando itens possuem a mesma demanda.

A. Métodos baseados em conteúdo

Esta abordagem utiliza como entrada informações descrevendo os itens e o perfil do usuário (figura 11). Baseado no conteúdo de cada item, verifica-se as características do item com as preferências do usuário (descritas pelo perfil do usuário) [43]. Por exemplo, através de uma função de pertinência pode-se recomendar ou não o item ao usuário.

A Filtragem Baseada em Conteúdo tem sido usada por várias aplicações para recuperar informações e gerar recomendações em sistemas de recuperação de informação [44], [50].

Para entendermos melhor o funcionamento dos métodos baseados em conteúdo, considere um perfil de usuário representado através de um vetor de palavras-chave. A cada palavra pertencente ao vetor foi atribuído um índice de importância. A importância de uma palavra k_i para um usuário p_j é determinada com alguma medida de peso $w_{i,j}$ que pode ser definida de diversas maneiras. Uma das maneiras existentes é através do cálculo de TF-IDF (*Term Frequency / Inverse Document Frequency*), segundo apresentado na equação 1.

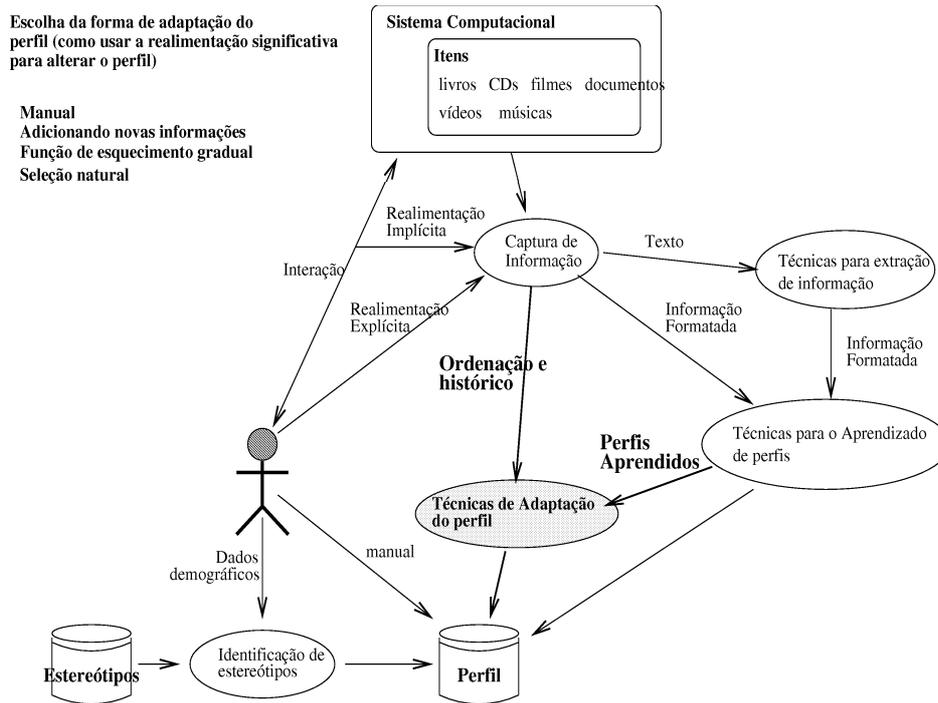


Figura 9. Formas de adaptação do perfil (baseado em [1])

$$w_{i,j} = f_{i,j} \times \log \frac{N}{n_i} \quad (1)$$

onde: $f_{i,j}$ é igual a frequência da palavra k_i no documento d_j . N é igual ao número total de documentos que o usuário selecionou como documento relevante e n_i é igual ao número de documentos em que a palavra k_i aparece. A frequência também pode ser normalizada, como apresentado na equação 2.

$$\frac{f_{i,j}}{\max_z f_{z,j}} \quad (2)$$

Em sistemas de recomendação baseados em conteúdo, a função de utilidade é usualmente definida como na equação 3 [1].

$$\mu(p, d) = \text{score}(\text{ContentBasedProfile}(p), \text{Content}(d)) \quad (3)$$

Tanto o $\text{ContentBasedProfile}(p)$ do usuário p como o $\text{Content}(d)$ do documento d podem ser representados como vetores $TF\text{-}IDF$ \vec{w}_p e \vec{w}_d de pesos de palavras-chave. Assim, a função de utilidade $\mu(p, d)$ pode ser definida como uma medida de similaridade utilizando o cosseno do ângulo de dois vetores (*cosine similarity measure*), por exemplo:

$$\mu(c, s) = \cos(\vec{w}_c, \vec{w}_s) = \frac{\sum_{i=1}^k w_{i,c} w_{i,s}}{\sqrt{\sum_{i=1}^k w_{i,c}^2} \sqrt{\sum_{i=1}^k w_{i,s}^2}} \quad (4)$$

Este é apenas um exemplo de técnica empregada na abordagem baseada em conteúdo. Pode-se empregar diversas outras técnicas nesta abordagem. A abordagem baseada em conteúdo

funciona muito bem com texto. Porém, em outros domínios é muito difícil extrair os atributos relevantes (i.e., imagens, áudio, vídeo). Dois itens distintos representados pelos mesmos conjuntos de atributos são iguais para um sistema baseado em conteúdo (artigo bem escrito versus artigo mal escrito). Os itens recomendados sempre serão iguais aos itens já conhecidos pelo usuário. Recomendação de itens muito similares (i.e., artigo de notícias descrevendo os mesmos eventos). A diversidade de recomendações geralmente é um atributo desejável em sistemas de recomendação. O usuário deve receber um espectro de opções e não um conjunto homogêneo de alternativas.

B. Métodos colaborativos

A abordagem colaborativa é usada quando podemos assumir que um usuário se comporta de forma semelhante à outros usuários, ou quando itens possuem a mesma demanda [20]. Este método tem sido usado com sucesso em domínios como de recomendação de filmes, músicas e jogos [43], [51].

A função de utilidade $\mu(p, o)$ do objeto o para o usuário p é estimada baseada na utilidade $\mu(p_j, o)$ descrita para o objeto o pelos usuários $p_j \in P$ que são similares ao usuário p (figura 12). Ou seja, um sistema de recomendação colaborativo tenta encontrar “pares” para o usuário p .

O valor de um objeto o para um usuário p , definido como $R_{p,o}$, é usualmente computado como um agregado (*aggr*) de valores de outros usuários (usualmente, os N mais similares) para o mesmo item o :

$$R_{p,o} = \text{aggr}_{p' \in \hat{P}} R_{p',o} \quad (5)$$

onde, \hat{P} representa o conjunto de N usuários que são os mais similares ao usuário p e que atribuíram um valor ao

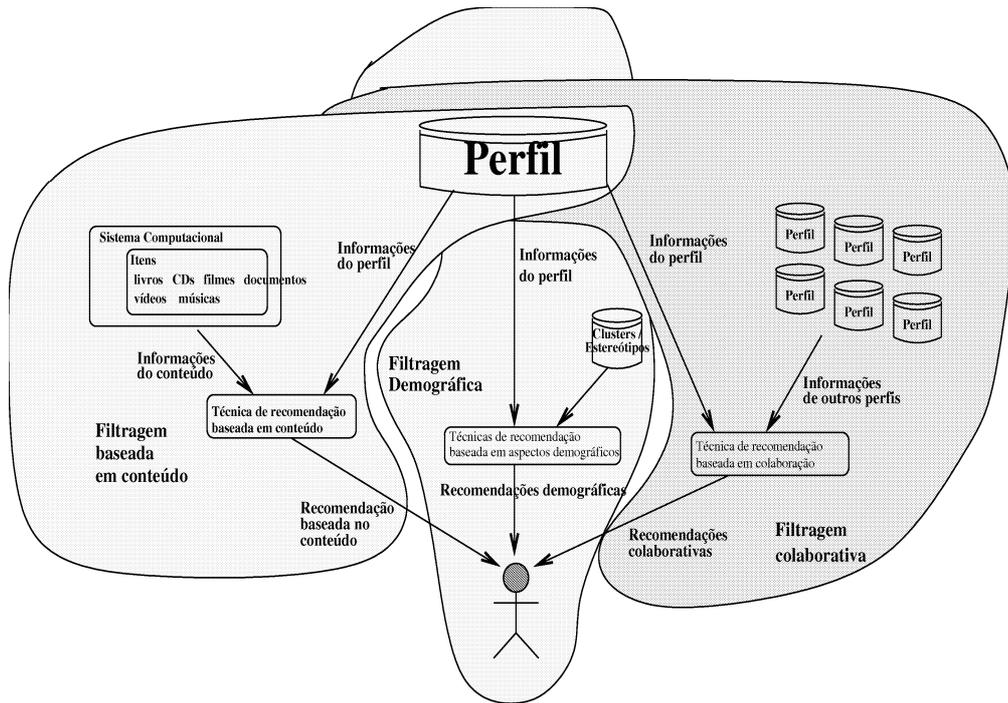


Figura 10. Exploração do perfil do usuário (baseado em [1])

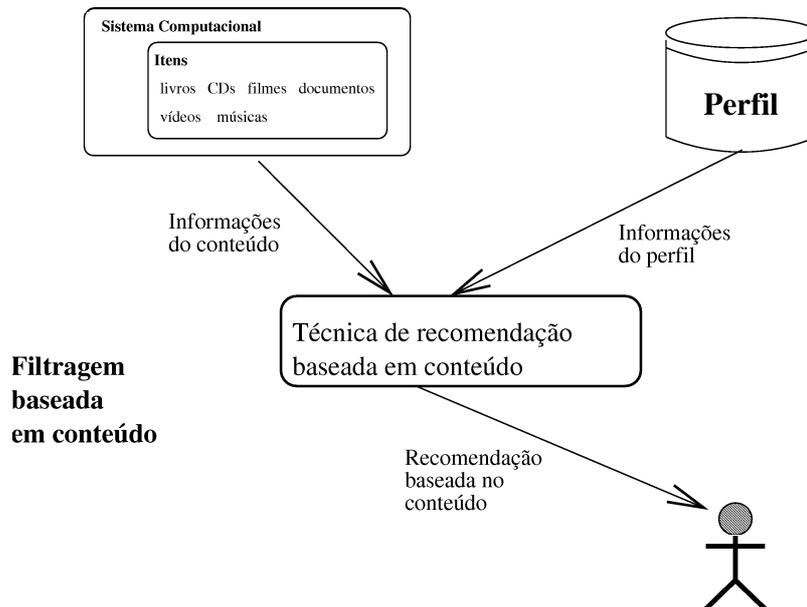


Figura 11. Filtragem baseada em conteúdo (baseado em [1])

objeto o . A função de agregação $R_{p,o}$ pode ser uma média simples:

$$R_{p,o} = \frac{1}{N} \sum_{p' \in \hat{P}} R_{p',o} \quad (6)$$

Ou uma função que leva em consideração a similaridade entre os usuários:

$$R_{p,o} = k \times \sum_{p' \in \hat{P}} sim(p, p') \times R_{p',o} \quad (7)$$

onde, k serve como um fator de normalização:

$$k = \frac{1}{\sum_{p' \in \hat{P}} sim(p, p')} \quad (8)$$

As duas abordagens mais comuns para calcular a similaridade entre usuários ($sim(x, y)$) são: o cálculo do coeficiente de correlação linear de *Pearson*⁶, e; o cálculo do cosseno do ângulo dos dois vetores.

⁶http://pt.wikipedia.org/wiki/Coeficiente_de_correlação_de_Pearson

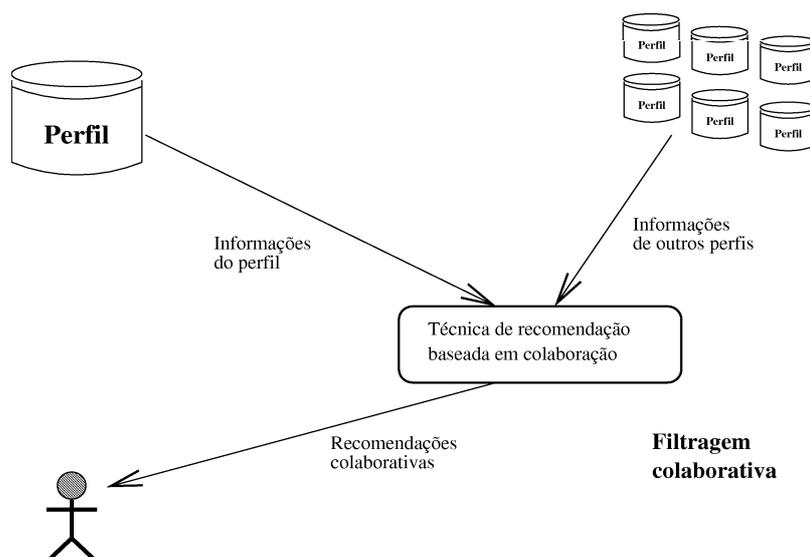


Figura 12. Filtragem colaborativa (baseado em [1])

Para que o sistema de recomendações possa gerar sugestões corretas, o sistema deverá primeiro aprender as preferências do usuário a partir de itens já mensurados pelo usuário - problema comum entre abordagens colaborativas e baseadas em conteúdo. Novos itens são adicionados regularmente nos sistemas de recomendação. No entanto, até que o item seja mensurado por um número substancial de usuários, o sistema não será capaz de recomendá-lo a nenhum outro usuário. Para este problema existem algumas heurísticas, por exemplo, filtragem demográfica.

C. Abordagens híbridas

Diversos sistemas de recomendação utilizam abordagens híbridas para combinar métodos baseados em conteúdo e colaborativos com o intuito de eliminar certas limitações dos sistemas baseados em conteúdo e dos colaborativos [7]. Existem diferentes maneiras de implementar um agente que utiliza uma abordagem híbrida. Entre elas, tem-se [1], [25]: (i) implementar métodos colaborativos e baseados em conteúdo separadamente e combinar as suas predições; (ii) incorporar algumas características da abordagem baseada em conteúdo na abordagem colaborativa; (iii) incorporar algumas características da abordagem colaborativa na abordagem baseada em conteúdo, e; (iv) desenvolver um modelo geral unificado que incorpora ambas as abordagens.

Ao combinar as predições pode-se diferenciar cenários: (i) combinar *rating* obtidos a partir de sistemas de recomendação individuais em uma recomendação final usando ou uma combinação linear ou um esquema de votação, e; (ii) possuir sistemas de recomendação individuais e escolher, com base em algum critério de qualidade, um sistema em um dado momento para realizar a recomendação. Os critérios de qualidade podem ser: (i) sistemas de recomendação com um maior nível de confiabilidade, e; (ii) sistemas de recomendação mais consistentes com os *ratings* passados do usuário.

As técnicas que incorporam algumas características da abordagem baseada em conteúdo na abordagem colaborativa são

baseadas nas tradicionais técnicas colaborativas, mas mantêm um perfil para cada usuário. Não apenas um $R_{p,o}$, mas um vetor de palavras-chave ou até mesmo informações pessoais [44]. Desta forma é possível tratar o problema dos usuários esparsos.

Também, um objeto pode ser recomendado a um usuário quando este objeto possui um alto grau de similaridade com o perfil do usuário. Não apenas quando este item é mensurado por algum usuário similar. Isto trata parcialmente o problema do item novo.

As técnicas que incorporam algumas características da abordagem colaborativa na abordagem baseada em conteúdo agrupam os perfis baseados em conteúdo. Diversos estudos empíricos comparando as abordagens híbridas com as abordagens puras tem mostrado que as abordagens híbridas possuem uma acurácia superior [52], [32].

V. CONSIDERAÇÕES FINAIS

Este tutorial descreveu e sintetizou os conceitos e técnicas utilizados na concepção de sistemas de recomendação que levam em consideração os interesses de usuários.

As técnicas e conceitos descritos foram divididos em dois grandes grupos: criação e manutenção do perfil do usuário, e; utilização do perfil do usuário. Na criação e manutenção do perfil do usuário os principais aspectos que devem considerados são: (i) a técnica de representação utilizada para armazenar o perfil do usuário; (ii) a técnica para a criação do perfil do usuário; (iii) a técnica de aprendizado do perfil do usuário; (iv) a fonte de realimentação relevante que representa os interesses do usuário, e; (v) a técnica de adaptação do perfil do usuário.

Sobre a utilização do perfil do usuário foram destacadas quatro abordagens: filtragem demográfica, filtragem baseada em conteúdo, filtragem colaborativa e abordagem híbrida.

Este trabalho é um *survey* teórico sobre o assunto. Para o desenvolvimento deste texto foram utilizadas as principais referências sobre o tema, em especial [25] e [1]. Trata-se de um texto voltado para pessoas que estão interessadas nos

fundamentos sobre modelagem de usuários e sistemas de recomendação. O texto tenta fornecer de forma ilustrativa esquemas que sintetizam os principais componentes utilizados na construção e utilização de perfis de usuários.

REFERÊNCIAS

- [1] M. Montaner, B. López, and J. L. de la Rosa, "A taxonomy of recommender agents on the internet," *Artificial Intelligence Review*, vol. 19, no. 4, pp. 285–330, June 2003.
- [2] E. Rich, "Users are individuals: Individualizing user models," *International Journal of Man-Machine Studies*, vol. 18, pp. 199–214, 1983.
- [3] —, "Users are individuals: Individualizing user models," *International Journal of Man-Machine Studies*, vol. 51, pp. 323–338, 1999.
- [4] H. Lieberman, Ed., *Your Wish is my Command*. Morgan Kaufmann Publishers, 2001.
- [5] A. Kobsa, "Generic user modeling systems," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-4, pp. 49–63, 2001.
- [6] C. Papatheodorou, "Machine learning in user modeling," in *Machine Learning and Applications*, G. Paliouras, V. Karkaletsis, and C. D. Spyropoulos, Eds. Berlin: Springer-Verlag Berlin Heidelberg, 2001, no. LNAI 2049, pp. 286–294.
- [7] C. Basu, H. Hirsh, and W. W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," in *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, Madison, Wisconsin, MW, 1998, pp. 714–720.
- [8] E. Bloedorn, I. Mani, and T. R. MacMillan, "Machine learning of user profiles: Representational issues," in *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Portland, Oregon, 1996, pp. 433–438.
- [9] B. Raskutti and A. Beitz, "Acquiring user preferences for information filtering in interactive multi-media services," in *Pacific Rim International Conference on Artificial Intelligence*, 1996, pp. 47–58.
- [10] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, March 1997.
- [11] G. Paliouras, C. Papatheodorou, V. Karkaletsis, and C. Spyropoulos, "Discovering user communities on the internet using unsupervised machine learning techniques," *Interacting with Computers*, vol. 12, pp. 761–791, March 2002.
- [12] G. I. Webb, M. J. Pazzani, and D. Billsus, "Machine learning for user modeling," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-4, pp. 19–29, 2001.
- [13] P. Langley, "User modeling in adaptive interfaces," in *Proceedings of the Workshop on Machine Learning in User Modeling. Advanced Course on Artificial Intelligence (ACAI'99)*, C. Papatheodorou, Ed. Greece: Springer, July 1999. [Online]. Available: <http://iit.demokritos.gr/skel/eetn/acai99/>
- [14] P. Brusilovsky and E. Schwarz, "User as student: Towards an adaptive interface for advanced web-based applications," in *Proceedings of 6th International Conference on User Modeling*, A. Jameson, C. Paris, and C. Tasso, Eds. Springer-Verlag, 1997, pp. 177–188.
- [15] D. N. Chin, "Knoome: modeling what the user knows," in *User models in dialog systems*, A. Kobsa and W. Wahlster, Eds. Springer-Verlag, 1989, pp. 74–107.
- [16] J. A. Self, "Bypassing the intractable problem of student modelling," in *Proceedings of the Intelligent Tutoring Systems Conference*, Montreal, 1988, pp. 107–123.
- [17] M. Bauer, W. Pohl, and G. Webb, Eds., *UM97 Workshop: Machine Learning for User Modeling*, Online proceedings: <http://www.dfki.uni-sb.de/~bauer/um-ws/>, 1997. [Online]. Available: <http://bistrica.usask.ca/UM/UM97/proceedings-overview.htm>
- [18] M. Bauer, P. J. Gmytrasiewicz, and W. Pohl, Eds., *UM99 Workshop: Machine Learning for User Modeling*, Online proceedings: <http://www.dfki.uni-sb.de/~bauer/um99-ws/>, 1999. [Online]. Available: <http://www.dfki.de/~bauer/um99-ws/>
- [19] J. Kay, Ed., *Proceedings of the Seventh International Conference on User Modeling*. <http://www.cs.usask.ca/UM99/papers.shtml>: Springer Wien New York, June 1999. [Online]. Available: <http://bistrica.usask.ca/UM/UM99/overview.htm>
- [20] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Distributed Systems OnLine*, vol. 1, January 2003. [Online]. Available: <file:///home/fabricio/doutorado/referencias/lin2003.htm>
- [21] P. Brusilovsky and C. Tasso, "Preface to special issue on user modeling for web information retrieval," *User Modeling and User-Adapted Interaction*, vol. 14, pp. 147–157, June 2004.
- [22] J. R. Quinlan, "Induction of decision trees," *Journal of Artificial Intelligence Research*, vol. 4, pp. 77–90, 1996.
- [23] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Elsevier, 2005.
- [24] L. Sterling and E. Shapiro, *The Art of Prolog*, 2nd ed. Cambridge, Massachusetts: The MIT Press, 1994.
- [25] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, June 2005.
- [26] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*. Cambridge University Press, 2008. [Online]. Available: <http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>
- [27] G. A. G. Lugo, "Um modelo de sistemas multiagentes para partilha de conhecimento utilizando redes sociais comunitárias," Ph.D. dissertation, Escola Politécnica da Universidade de São Paulo, Abril 2004.
- [28] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on web usage mining," *Communications of the ACM*, vol. 43, no. 8, pp. 142–151, 2000.
- [29] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
- [30] S. Mewggleton and L. D. Raedt, "Inductive logic programming: Theory and methods," *Journal of Logic Programming*, vol. 19, no. 20, pp. 629–679, 1994.
- [31] P. Baffes and R. Mooney, "Refinement-based student modeling and automated bug library construction," *Journal of Artificial Intelligence in Education*, vol. 7, pp. 75–116, 1996.
- [32] D. Billsus and M. J. Pazzani, "A hybrid user model for news story classification," in *Proceedings of the Seventh International Conference on User Modeling*, J. Kay, Ed. Banff, Canada: Springer Wien New York, June 1999, pp. 99–108. [Online]. Available: <http://bistrica.usask.ca/UM/UM99/overview.htm>
- [33] R. B. Segal and J. O. Kephart, "Mailcat: an intelligent assistant for organizing e-mail," in *AGENTS '99: Proceedings of the third annual conference on Autonomous Agents*. New York, NY, USA: ACM Press, 1999, pp. 276–282.
- [34] H. Lieberman, "Letizia: An agent that assists web browsing," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Montreal, Canada, 1995, pp. 924–929. [Online]. Available: <http://lieber.www.media.mit.edu/people/lieber/Lieberary/Letizia/Letizia.html>
- [35] K. Kwok and M. Chan, "Improving two-stage ad-hoc retrieval for short queries," in *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, 1998, pp. 250–256.
- [36] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell, "Learning to classify text from labeled and unlabeled documents," in *Proceedings of the 15th International Conference on Artificial Intelligence*, Madison, Wisc, 1998, pp. 792–799.
- [37] Z. Chen and B. Zhu, "Some formal analysis of rocchio's similarity-based relevance feedback algorithm," in *Lecture Notes in Computer Science*, no. 1969, 2000, pp. 108–119.
- [38] A. Moukas, "Amalthaea: Information discovery and filtering using a multiagent evolving ecosystem," *Applied Artificial Intelligence: An International Journal*, vol. 11, no. 5, pp. 437–457, 1997.
- [39] A. Rashid, I. Albert, D. Cosley, S. Lam, S. McNee, J. Konstan, and J. Riedl, "Getting to know you: Learning new user preferences in recommender systems," in *Proceedings of the 2002 International Conference on Intelligent User Interfaces*, San Francisco, CA, 2002, pp. 127–134.
- [40] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 2002, pp. 253–260.
- [41] S. E. Middleton, "Capturing knowledge of user preferences with recommender systems," Ph.D. dissertation, University of Southampton, May 2003.
- [42] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Machine Learning*, vol. 23, pp. 69–101, 1996.
- [43] I. Koychev and I. Schwab, "Adaptation to drifting user's interests," in *Proceedings of ECML 2000 Workshop: Machine Learning in New Information Age*, 2000.
- [44] M. Pazzani and D. Billsus, "Learning and revising user profiles: The identification of interesting web sites," *Machine Learning*, vol. 27, pp. 313–331, 1997.

- [45] K. Lang, "Newsweeder: Learning to filter news," in *Proceedings of the Twelfth International Conference on Machine Learning*, Lake Tahoe, CA, 1995, pp. 331–339.
- [46] R. Mooney, P. Bennet, and L. Roy, "Book recommending using text categorization with extracted information," in *AAAI/ICML-98 Workshop on Learning for Text Categorization*, Madison, Wisc, 1998.
- [47] R. Klinkenberg and I. Renz, "Adaptive information filtering: learning in the presence of concept drift," in *AAAI/ICML-98 Workshop on Learning for Text Categorization. Technical Report WS-98-05*, Madison, Wisc, 1998.
- [48] M. C. Rosatelli and P. A. Tedesco, "Diagnosticando o usuário para criação de sistemas personalizáveis," in *Anais do XXIII Congresso da Sociedade Brasileira de Computação*, vol. VIII, 2003, pp. 153–201.
- [49] P. Chiu and G. Webb, "Using decision trees for agent modeling: improving prediction performance," *User Modeling and User-Adapted Interaction*, vol. 8, pp. 131–152, 1998.
- [50] F. Esposito, D. Malerba, G. Semeraro, N. Fanizzi, and S. Ferilli, "Adding machine learning and knowledge intensive techniques to a digital library service," *International Journal on Digital Libraries*, vol. 2, no. 1, pp. 3–19, October 1998.
- [51] M. Balabanovic and Y. Shoham, "Fab: content-based, collaborative recommendation," *Communications of the ACM*, vol. 40, no. 3, pp. 66–72, March 1997.
- [52] T. Tran and R. Cohen, "Hybrid recommender systems for electronic commerce," in *Proceedings of AAAI00 workshop on Knowledge-Based Eletronic Markets*, 2000.



Fabrício J. Barth Possui graduação em Ciências da Computação pela Fundação Universidade Regional de Blumenau (2000), mestrado em Engenharia Elétrica pela Universidade de São Paulo (2003) e doutorado em Engenharia Elétrica pela Universidade de São Paulo (2009). Atualmente é responsável pela área de Recuperação de Informação da empresa Apontador (www.apontador.com.br). Tem experiência na área de Ciência da Computação, atuando principalmente nos seguintes temas: recuperação de informação, mineração de

textos, mineração de dados, recuperação e mineração de informações na web, sistemas adaptativos e ferramentas para apoio à Gestão de Conhecimento. Pode ser encontrado nos endereços fabricao.barth@gmail.com e <http://fbarth.net.br>